Artificial intelligence-augmented, label-free molecular imaging method for tissue identification, cancer diagnosis, and cancer margin detection

JIASONG LI,[1,6] JUN LIU,[1,2,6] YE WANG,[1,2,6] YUNJIE HE,[1] KAI LIU,[1] RAKSHA RAGHUNATHAN,[1] STEVEN S. SHEN,[3] TIANCHENG HE,[1] XIAOHUI YU,[1] REBECCA DANFORTH,[1] FEIBI ZHENG,[4] HONG ZHAO,[1] AND STEPHEN T. C. WONG[1,3,5,*]

[1]*Department of Systems Medicine and Bioengineering, Houston Methodist Cancer Center, Weill Cornell Medicine, Houston, TX 77030, USA*

[2]*Department of Breast-thyroid-vascular Surgery, Shanghai General Hospital, Shanghai Jiao Tong University, 201620, Shanghai, China*

[3]*Department of Pathology and Genomic Medicine, Houston Methodist Hospital, Weill Cornell Medicine, Houston, TX 77030, USA*

[4]*Department of Surgery, Houston Methodist Hospital, Weill Cornell Medicine, Houston, TX 77030, USA*

[5]*Department of Radiology, Houston Methodist Hospital, Weill Cornell Medicine, Houston, TX 77030, USA*

[6]*These authors contributed equally to this work*

[*]*stwong@houstonmethodist.org*

Abstract: Label-free high-resolution molecular and cellular imaging strategies for intraoperative use are much needed, but not yet available. To fill this void, we developed an artificial intelligence-augmented molecular vibrational imaging method that integrates label-free and subcellular-resolution coherent anti-stokes Raman scattering (CARS) imaging with real-time quantitative image analysis via deep learning (artificial intelligence-augmented CARS or iCARS). The aim of this study was to evaluate the capability of the iCARS system to identify and differentiate the parathyroid gland and recurrent laryngeal nerve (RLN) from surrounding tissues and detect cancer margins. This goal was successfully met.

1. Introduction

The incidence of thyroid cancer has tripled in the past three decades. The American Cancer Society's estimates in the United States for 2019 are 52,070 new cases of thyroid cancer (14,260 in men and 37,810 in women) and 2,170 deaths from thyroid cancer (1,020 men and 1,150 women). Increased incidence of thyroid cancer has led to an increase in thyroidectomies in the United States. The total number of thyroidectomies increased by 39% (from 66,864 to 92,931 cases) annually in the United States from 1996 to 2006. Outpatient procedures increased drastically by 61% while inpatient procedures increased by 30%. The aggregate national charges for inpatient thyroidectomy in the United States alone increased from $464 million in 1996 to $1.37 billion in 2006 [1].

Hypoparathyroidism is the most common surgical complication associated with thyroid surgeries [2]. Due to the differences in the standards of postoperative recovery of parathyroid function, and varying biochemical assays used across institutions, incidence of surgical hypoparathyroidism varies from 0.33% to 68%. Hypoparathyroidism occurs during neck surgeries after inadvertent trauma, devascularization, or removal of parathyroid glands. Its risk increases when a central lymph node dissection is performed during thyroid cancer surgery. Its risk can be as high as 30% after repeat thyroid, parathyroid, or anterior neck surgeries [2]. Multiple imaging techniques

have been developed to localize the abnormal parathyroid gland including high-resolution ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and nuclear medicine sestamibi scan (MIBI) [3]. However, positive predictive localization values of these methods vary widely from 36-100% as they are limited by their spatial resolutions, protocols for use, and the experience of the operators and radiologists. Furthermore, these preoperative techniques are only capable of providing a rough estimate of localization of the hyperplasia of parathyroid glands. Several techniques have been used to assist intraoperative localization. But all of them have their own limitations. Intraoperative ultrasound is not useful if parathyroid adenoma was not visible on preoperative ultrasound. The intraoperative parathyroid hormone assay (IOPTH) is capable of assessing the adequacy of resection by measuring the IOPTH levels after removing all hyperfunctioning parathyroid glands [4,5]. However, it can neither detect parathyroid glands nor differentiate abnormal parathyroid glands from normal glands. Another challenge is parathyroid hyperplasia, where the gland mass is smaller than that of parathyroid adenoma [6].

During surgeries, injuries to the parathyroid gland and recurrent laryngeal nerve (RLN) should be avoided as they can significantly decrease the patient's quality of life [7]. Parathyroid gland injury can induce transient or persistent hypoparathyroidism, which manifests as paresthesia, tetany, and seizures while RLN injury can cause hoarseness or vocal cord paralysis [8]. The reported incidental rate of parathyroidectomy ranges between 6.4%−31% during thyroid surgery, and the rate of persistent vocal cord paralysis is about 1% [9,10]. Hence, maintaining well perfused parathyroid glands and proper identification of the RLN are essential during thyroid carcinoma surgery. However, misidentification of the parathyroid glands as lymph nodes, adipose tissues or extra-thyroidal thyroid tissues [11] and RLN as surrounding fibrous connective tissues is very common. These misidentifications are due to similarities in appearance, especially during repeated operative neck surgeries [12]. Techniques like nanoparticle injections and intraoperative nerve monitoring devices have been developed to address these issues [13–16]. However, all these techniques have limitations including indirect identification. Hence, the efforts to reduce the occurrence of these two severe side effects of thyroid surgery remain unfruitful.

A more sensitive method to identify parathyroid glands and RLN during thyroid surgery is highly desirable [17]. Four methods currently under investigation to identify the parathyroids during thyroid surgery include parathyroid washout test, fluorescence imaging, auto-fluorescence imaging, and label-free parathyroid morphology [18–23]. Through the function test, Bian et al. demonstrated that the rapid intraoperative parathyroid hormone (rIO-PTH) assays using fine needle aspiration (FNA) is a plausible method for identifying parathyroid tissue [19]. However, FNA may damage parathyroid glands, and the rIO-PTH test takes 10-20 minutes or longer to complete [19]. Through intravenous injection of fluorophores, Hyun et al. and Suh et al. showed that a fluorescence imaging system can identify parathyroid and thyroid glands with high-sensitivity and unambiguous mode [20,21]. However, the suitability of the fluorophores for clinical use needs evaluation and requires clinical trials for FDA approval as exogenous agents. Interestingly, Tummers et al. introduced near-infrared fluorescence imaging during parathyroid surgery and found that administration of low dose methylene blue (which is already used in clinics to guide sentinel lymph node biopsy in breast cancer surgery) could identify parathyroid adenomas and normal parathyroid glands [24]. Nonetheless, since the side effects of intravenous injection of fluorophores have not been determined, auto-fluorescence imaging is preferred to detect parathyroid glands instead. A fluorescence detection apparatus used by Shinden et al. showed that auto-fluorescence detected by near-infrared light could identify the parathyroid gland during thyroid surgery [22]. McWade et al. used near-infrared light to detect the auto-fluorescence of parathyroid and surrounding tissues and found that the parathyroid gland's fluorescence was stronger than that of thyroid while surrounding tissues showed no auto-fluorescence [23,25,26]. Falco et al. found that the mean intensity of parathyroid gland fluorescence using near-infrared light was significantly higher than the intensity of the thyroid

gland and the background [27]. They also showed that the number of parathyroid glands identified increased dramatically due to the use of near-infrared light for parathyroid gland visualization [27]. However, auto-fluorescence is an indirect method and is restricted in routine clinical use as its signal intensity fluctuates greatly and interferes with surrounding tissues. More accurate and straightforward methods to identify the parathyroid by morphology *ex vivo* have thus been explored. Conti de Freitas et al. and Hou et al. independently found that optical coherence tomography (OCT) can reliably identify the thyroid, parathyroid, adipose, muscle, and lymph nodes [18,28]. Nevertheless, the low resolution (a few microns) and low contrast of OCT images when compared to images acquired with hematoxylin and eosin stain (H&E stain) make it unsuitable for detailed investigations. There are FDA approved devices available to detect parathyroid glands, but they make use of the aforementioned modalities with limitations. For example, the Fluobeam 800 system is an autofluorescence system in which the signal intensity still varies and the detection sensitivity is poor, whereas the PTeye system can only provide a localized signal at a focal-spot sized location and cannot show the entire parathyroid gland. As far as RLN identification is concerned, no approved method exists besides restrictive intraoperative nerve monitoring.

Coherent anti-Stokes Raman scattering (CARS) imaging is a label-free, 3-dimensional (3D), real-time, nonlinear optical imaging technique based on the intrinsic molecular vibrations of the sample (thus also known as molecular vibration imaging). CARS imaging provides superior spatial resolution (down to the nm level) [29]. Since it is dependent on the intrinsic properties of the sample, no external contrast agents or dyes are required. CARS is capable of imaging the tissue morphology as well as providing subcellular information, both of which are required for pathological analyses. Applications of CARS in medicine have been increasing recently, including several pre-clinical studies on lungs, brain, breast, kidney, prostate, and skin cancer [29–34]. These applications mainly focus on differentiating carcinoma from benign tissue and identifying margins of the cancerous tissue.

However, the identification and differentiation of various tissue types within a complex surgical region using high resolution microscopy images are challenging. It also challenges conventional image segmentation and classification methods as it requires considerable parameter tuning or human intervention. Deep learning networks are being used to help with these issues. These networks learn from labeled training data fed by human experts and provide fast and improved classification and pattern recognition. This decreases the cost of model construction and shortens the time between actions. Using deep learning for biological samples relies on deep convolutional neural networks (DCNN) [35,36]. Inception-v3 and VGG16 are both DCNN models that achieved 21.2% and 24.4% top-1 error on ILSVRC 2012 classification challenge dataset [37,38]. Their layered weights were implemented in Keras (based on Tensorflow, Google). Considering the advantages of transfer learning, such as decreasing the lower limit of training dataset volume, reducing convergent time, and improving model performance, the pre-trained weighted layers of the convolutional parts of Inception-v3 and VGG16 were added to our image quantitation models [39–42]. Another DCNN model, CNN4, that does not include pre-trained weights from any model, was constructed as a control against those two pre-trained models mentioned above. The entire system (imaging + analysis) used in this study was called an artificial intelligence-augmented CARS (iCARS) imaging system.

We evaluated the capability of the iCARS imaging system to identify porcine parathyroid glands, RLN, and surrounding tissues, including thyroid, lymph node, fat tissues, and muscle. We also acquired a high-resolution 3D image of the murine thyroid gland *in situ* to show the potential use of this system as an intraoperative guiding tool. The deep-learning algorithm that was developed, helped with instant identification of all porcine tissue types and helped differentiate between them. Finally, we also used the iCARS imaging system to identify human thyroid cancer tissue and detect cancer margins. Results showed that the iCARS system was

successful in automatic and rapid real-time identification of different types of porcine tissues, and human thyroid cancer tissue while detecting the cancer margins.

2.   Materials and methods

2.1.   Tissue sample preparation

For thyroid cancer imaging, human thyroid tissues (both cancerous and healthy) were obtained from patients undergoing thyroid/parathyroid surgery at Houston Methodist Hospital, Houston, Texas, USA and Shanghai General Hospital, Shanghai, P.R. China, following an institutional review board approval. The excised tissue samples were cut into 5 mm chunks and then immediately snap-frozen in liquid nitrogen for storage. Frozen tissue samples were passively thawed for 30 minutes at room temperature before iCARS imaging.

Fresh porcine esophagi with surrounding tissues were dissected immediately after sacrifice and shipped overnight (Loretta Tomlin, Tyler, TX, USA). The thyroid glands, parathyroid glands, lymph nodes, fat tissues, RLN, and muscle tissues were harvested by an experienced thyroid surgeon. All tissues were cut into two parts and kept in phosphate-buffered saline. One piece was placed on a slide immediately for iCARS imaging, and the other part was fixed with paraformaldehyde for H&E staining. Human H&E slides of corresponding tissues were obtained from Department of Pathology, Houston Methodist Hospital (Houston, TX, USA) for comparisons. H&E slides of the porcine tissues were read by an experienced pathologist to confirm the tissue type. These were then compared to the iCARS images of the porcine tissues and the corresponding H&E slides of human tissue.

To test the potential of iCARS imaging as an intraoperative guiding tool, the murine thyroid was imaged in situ. A healthy mouse (10 months old) was euthanized by CO_2 inhalation and its neck was cut open to mimic thyroid/parathyroid surgery. A cover glass was placed over the field of interest (determined by an experienced surgeon) to ensure a flat surface for imaging. Within 10 minutes of euthanasia, the animal was prepared, and the tissue was imaged.

2.2.   Principle of CARS

CARS is sensitive to the same vibrational signatures of molecules as Raman spectroscopy, typically the nuclear vibrations of chemical bonds. However, unlike the case of Raman scattering, CARS requires two short-pulse beams (a pump and a Stokes beam) to irradiate a molecule (sample). The frequency difference between these two beams is tuned to be equal to the vibration frequency of the molecule. When this condition is met, coherently vibrating molecules in the focal point of the pump and Stokes beam scatter the signal, and the obtained signal has a much higher frequency and much higher intensity than a signal from regular Raman scattering. Thus, compared to Raman imaging, CARS offers much higher sensitivity.

2.3.   Label-free optical imaging system

The optical source of our iCARS system is a dual-output ultrafast tunable laser (Insight Deepsee, Spectra-Physics, Santa Clara, CA, USA) that provides 120 fs, 80 MHz pulses. The output of 1040 nm was utilized as the Stokes beam, and the tunable output (range from 680nm to 1300nm) serves as the pump beam. The Stokes beam and the pump beam are overlapped both temporally and spatially by adjusting a time-delay line (DL) in the temporal domain and using a long-pass dichroic mirror (Chroma, VT) in the spatial domain to generate a steady CARS signal. A dichroic mirror was used to separate emission signals from excitation laser beams. The upright scanning microscope is modified from a Nikon Eclipse FN1 fixed stage microscope using a 2D galvanometer. The galvanometer mirrors are controlled by the image acquisition software described in section 2.4. A 1.2-NA water immersion objective (60X, Olympus) was used for sample imaging. A red-light-sensitive photomultiplier tube (PMT, R3896, Hamamatsu, Japan)

with a spectral response of 185 nm to 900 nm was used as the detector. The CARS signal was separated from the remaining using a bandpass filter in the setup. Figure 1 shows a picture of the CARS system used.
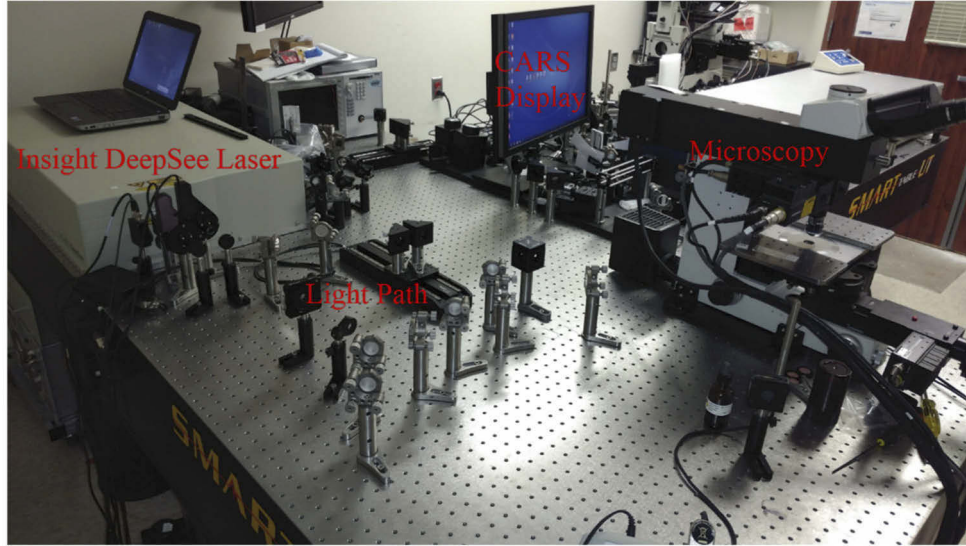

Fig. 1. The femtosecond CARS system used for this study.

2.4. Image acquisition

The tissue samples were placed on a 170 μm cover slide (VWR, Radnor, PA, USA) and then inverted on an imaging chamber to avoid possible compression. For iCARS imaging, the pump beam was tuned to 802 nm, and the Stokes beam was fixed at 1040 nm to probe the symmetric stretching frequency of CH_2 bond at 2845 cm^{-1} [43]. iCARS signals were generated at 663 nm. The image (512×512 pixels) was acquired and displayed in real-time using the ThorImage 3.0 software (Thorlabs, Inc), and the average power was about 75 mW for pump beam and 35 mW for Stokes beam. Bright-field images of the H&E slides were examined with an Olympus BX51 microscope as a standard control.

2.5. Image analysis

To statistically analyze the cell signals distribution of parathyroid and thyroid gland in CARS images, cell image segmentation and analyses were performed. We first segment the cellular areas from the original images. Due to the high resolution of our CARS images, we simply applied a previously proposed segmentation method [44,45]. In our application, the 2D scalar function $\varphi(x, y, t)$ was seen as a distance map such that each pixel (x, y) in image I receives a value of the $\varphi(x, y, t)$ function. Pixels that lie on the contour receive a value of 0 while the remaining pixels receive values based on their distance from the curve. Since the evolving front $\gamma(t)$ [46] is a level set of the scalar function φ for every time t, the evolving contour obtained can be expressed as:

$$\theta(\gamma(t), t) = 0 \qquad (2.1)$$

Derivation with respect to t, using the chain rule yielded

$$\theta_t + \nabla_\theta \cdot \gamma_t = 0 \qquad (2.2)$$

Let σ be the speed in which the contour propagates in the direction normal to the curve. Hence

$$\sigma = \gamma_t \cdot \rho \tag{2.3}$$

where

$$\rho = \frac{\nabla_\theta}{|\nabla_\theta|} \tag{2.4}$$

Therefore Eq. (2.2) becomes:

$$\theta_t + \sigma |\nabla_\theta| = 0 \tag{2.5}$$

We get a partial differential equation (PDE) on θ with an initial curve $\theta(\gamma(t), t = 0)$. This equation can be solved using finite difference approximations for the special and temporal derivatives [47]. To account for regularization, a curvature dependent speed $\sigma = \sigma(\text{n})$ [47] is used with a constant term σ_0:

$$\sigma(\text{n}) = \sigma_0 - n \tag{2.6}$$

The edge-stop term [47] is defined by the function:

$$g(x, y) = \frac{1}{1 + |\nabla G_\sigma * I(x, y)|} \tag{2.7}$$

This function behaves as an inverse of a gradient detector. It, therefore, approaches zero in the vicinity of an edge, bringing the velocity to a stop. So (2.5) can be switched as below:

$$\theta_t + g(x, y)(\sigma_0 - n)|\nabla_\theta| = 0 \tag{2.8}$$

Incorporating the classical equation for the curvature n:

$$n = div\left(\frac{\nabla_\theta}{|\nabla_\theta|}\right) = \frac{\theta_{yy}\theta_x^2 - 2\theta_x\theta_y\theta_{xy} + \theta_{xx}\theta_y^2}{\left(\theta_x^2 + \theta_y^2\right)^{3/2}} \tag{2.9}$$

and changing the σ_0, we get the final Level Set flow as a PDE:

$$\theta_t = g(x, y)|\nabla_\theta|\left(\sigma_0 + div\left(\frac{\nabla_\theta}{|\nabla_\theta|}\right)\right) \tag{2.10}$$

The final equation we use in our algorithm is:

$$\theta_t = g(x, y)|\nabla_\theta|\sigma_0 + g(x, y)|\nabla_\theta|div\left(\frac{\nabla_\theta}{|\nabla_\theta|}\right) \tag{2.11}$$

By solving the PDE following the approach [47], we can segment the target cell or tissues by using the final curve θ_t. After the target cells or tissues are segmented, the intensity distribution can be computed over the segmented area of parathyroid, thyroid, lymph nodes and muscle tissues. In our application, we defined the intensity distribution (i.e. the mean and standard deviation) in the target cells as the feature to identify the differences among the different tissues. After obtaining the segmented CARS images, the manual tissue identification was performed. This entire analysis was implemented using MATLAB 2020a on our server (Intel Xeon Gold 5122 CPU @ 3.60GHz with Ram 123GB).

2.6.   *Tissue identification and differentiation*

Two methods were used for tissue identification and differentiation. The first was a manual method that used normalized intensities of CARS images to differentiate between the tissues. The second method made use of deep learning algorithms in real-time, thus reducing the time for decision making. For tissue identification (parathyroid, thyroid, lymph node, fat, nerve, and muscle), the regular retrained inception-v3 algorithm was used. Since this algorithm achieved an accuracy of only 80.21%, an improved Bayesian retrained Inception-v3 algorithm was then applied. For automatic cancer margin detection, the U-Net structure was used. All these methods and their results are described in detail in the results section.

3. Results

The first part of the results section discusses the CARS images acquired and the manual quantifications that were performed to differentiate between the different types of tissue (section 3.1). The second half of the results section discusses the application of the iCARS system (section 3.2 onwards). Section 3.2 includes a description of deep learning methods developed as a part of this study and their results when applied to our CARS images.

3.1. CARS imaging: manual tissue identification and differentiation

3.1.1. Identification of thyroid, parathyroid glands, and surrounding tissue
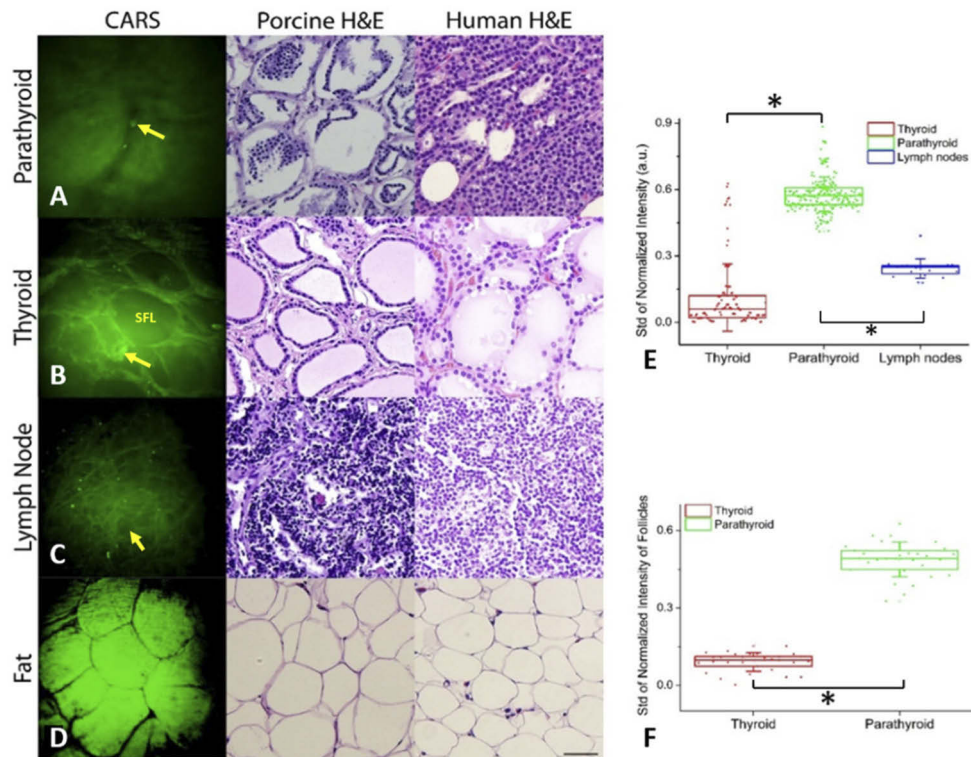

Fig. 2. A: CARS and histological images of porcine and human parathyroid tissue. The yellow arrow indicates the fat droplets present that help in differentiating the parathyroid from other tissues. B: CARS and histological images of porcine and human thyroid tissue. SFL refers to the spherical follicular structures while the yellow arrow indicates the darker nuclei. C: CARS and histological images of porcine and human lymph nodes. Yellow arrow indicates the cell clusters. D: CARS and histological images of porcine and human fat tissue. Scale bar for all images: 50 μm. E: The box plot shows the mean (horizontal line in the box) and standard deviation (whiskers) of the normalized intensity of CARS images of thyroid, parathyroid, and lymph nodes. The dots in each of the plots are the data points, depicting the data spread. The asterisk indicates a statistical significance of P<0.01 using an unpaired student's t-test. F: The box plot shows the mean (horizontal line in the box) and standard deviation (whiskers) of normalized intensity of randomly selected follicles from CARS images of thyroid and parathyroid tissues. The dots in each plot show the data points depicting the data spread. The asterisk indicates a statistical significance of P<0.01 using a paired student's t-test.

The CARS (porcine) and histological (porcine and human) images of the parathyroid are shown in Fig. 2(A). Normal parathyroid images have well-defined cytoplasmic membrane and fat droplets, which are helpful in differentiating parathyroid from thyroid. In the CARS image of the porcine parathyroid gland, we noticed multiple pool-like structures surrounded or partially filled by chief cells. Adipocytes were also seen in the CARS images of parathyroid, helping to differentiate the parathyroid gland from thyroid or lymph node tissues.

Histological images of human and porcine thyroid glands (Fig. 2(B)) showed morphological features such as spherical follicular structures with homogeneous internal colloids made up mostly of thyroglobulin and surrounded by a layer of follicular cells. Similar spherical follicular structures were identified in the CARS images of porcine thyroid tissue. They presented themselves in oval shapes and were filled with a homogeneous medium. The surrounding basement membrane was slightly dimmer while the nuclei of the follicular cells were much darker. The cytoplasm of the follicular cells was much brighter than its other constituents in the CARS image. This could be because of the presence of certain CH_2-rich subcellular structures in them, such as lipid drops. Generally, a typical CARS image of porcine thyroid gland appears as pool-like follicles with uniform intensity inside the follicles.

Lymph nodes generally consist of cortex, paracortex, and medulla. Figure 2(C) shows the histological and CARS images of lymph nodes. In the histological images, both porcine and human lymph nodes were rich in cells, with the nuclei occupying more than half the cell. CARS images of porcine lymph node medulla showed that the nuclei's shape and size varied, gathering (dimmer signal) and embedding in surrounding net-like fibrous connective tissues (brighter signal), similar to H&E images. Cell clusters (gathered cells) were seen all over the lymph node images while such cell distribution was only found inside the follicles in the parathyroid gland images. No similar clusters were found in thyroid images. Unlike the parathyroid and thyroid images, there were no pool-like follicles in the lymph node images. Last but not least, the CARS images of the lymph node did not show any fat droplet, which was seen in the parathyroid gland images.

Fat tissues (Fig. 2(D)), generally, are an accumulation of oval, and size-varied adipocytes, in which cytoplasm is uniform with a well-defined boundary surrounded by a fine network of reticular fibers. Since our CARS system was tuned to be sensitive to the chemical bonds of CH_2, which is abundant in fat tissues, these tissues showed up brightly with a thin, dimmer, and dedicated cell membrane in CARS images.

In each CARS image, the intensity of every pixel was normalized to the mean intensity value of that image. Then the standard deviation of each normalized pixel was calculated with respect to the mean of all the normalized values of that image. Thus, each image was represented by only one standard deviation value. Figure 2(E) shows a comparison of these values for the parathyroid, thyroid, and lymph nodes. A total of 107 thyroid images, 207 parathyroid gland images, and 20 lymph node images were analyzed. Figure 2(E) showed that the thyroid had the smallest values compared to the parathyroid and lymph nodes. This was due to the presence of minimal fat in thyroid tissues, which led to a relatively uniform intensity inside each follicle. On the other hand, values of the lymph nodes were greater than that of the thyroid tissue due to the high contrast between the bright cell membrane and the dark cell nuclei in the lymph nodes. This was also because of the cell clusters. The values of the parathyroid images were the greatest due to the high concentration of fat, as well as the cell clusters inside the follicles. An unpaired student's t-test was performed between the standard deviation values of parathyroid images and lymph node images, and parathyroid images and thyroid images, both resulted in $P<0.01$, indicating that parathyroid was successfully differentiated from lymph nodes and thyroid.

Even though the parathyroid tissue was differentiated from thyroid successfully by analyzing the intensity of the whole CARS image, a small overlap in data was noticed (Fig. 2(E)). Therefore, further data processing to distinguish parathyroid from thyroid was conducted. As mentioned

before, the structure of the thyroid follicle was uniform, whereas the structure of the parathyroid follicle was a cluster of cells. Thus, the focus was now shifted to the standard deviation of the normalized intensity of each follicle alone instead of the entire image. Thirty follicles were randomly selected for analysis. Since the structure of the thyroid follicle is much more uniform than the parathyroid follicle, the values of standard deviation of normalized intensity of thyroid follicles were much smaller than that of parathyroid follicles, as seen in Fig. 2(F). A paired student's t-test was performed and a statistical significance of P< 0.01 was seen again. However, this time, there was no data overlap. Hence the parathyroid images were successfully differentiated.

Apart from the manual quantifications for these tissues, a blind observer study of both the H&E images and CARS images were performed. The results from the reading of the H&E images (which was performed by an experienced pathologist) was considered the gold standard. This data was used as the basis for the deep learning algorithms. Table 1 shows the confusion matrix for the identification of thyroid, parathyroid, and lymph node. Every time the CARS reading matched the H&E readings (gold standard), the corresponding location on the confusion matrix was filled up. For example, the readings of the CARS images of the thyroid tissue matched with the corresponding H&E image reading 101 times. The overall accuracy of the classification was calculated to be 98.8%. The accuracy, precision, sensitivity, specificity, and F-score of thyroid identification were calculated to be 94.4%, 99.4%, 98.1%, 99.1%, and 96.2%, respectively. The accuracy, precision, sensitivity, specificity, and F-score of parathyroid identification were calculated to be 99.0%, 97.2%, 99.0%, 95.3%, and 98.1%, respectively.

Table 1. The confusion matrix for the identification of thyroid, parathyroid, and lymph node (total of 334 images).

		H&E		
		Thyroid	Parathyroid	Lymph node
CARS	Thyroid	101	2	0
	Parathyroid	6	205	0
	Lymph node	0	0	20

3.1.2.   RLN and muscle imaging and identification

Figures 3(A) and (B) show the images of (porcine and human) RLN and muscle respectively. In the H&E images of porcine and human RLN (Fig. 3(A)), straps with less Schwann's nuclei surrounded by smooth muscles can be identified directly. In the CARS image of the porcine RLN (Fig. 3(A)), these straps showed up with higher intensity because nerve bundles contain rich lipid membranes. The dark areas (depicted in the image) around the high intensity straps are the nuclei of Schawann's cells. Uniformly bright and relatively wide parallel straps seen in the CARS image are the typical characteristics of RLN. Therefore, the RLN can be successfully identified by CARS imaging.

Bundles of skeletal muscle fibers were seen in the H&E images of transverse sections of porcine and human muscle (Fig. 3(B)). In the CARS image of porcine muscle (longitudinal section) (Fig. 3(B)), myocytes appeared in longitudinal narrow and elongated fiber-like bands, unlike the RLN which appeared as long straps separated by darker regions.

Since our CARS imaging system is tuned to be sensitive to CH_2 bonds, rich lipid regions appeared brighter in the images. Hence, RLN is characterized by bright straps with a few darker intervals whereas muscle showed up as orderly bands with mostly uniform intensity. Intensity analysis similar to what was described earlier in the case of the parathyroid identification was performed to distinguish between the CARS images of RLN and muscle. These results are shown in Fig. 3(C). The box plots show the mean and standard deviation of the values (standard
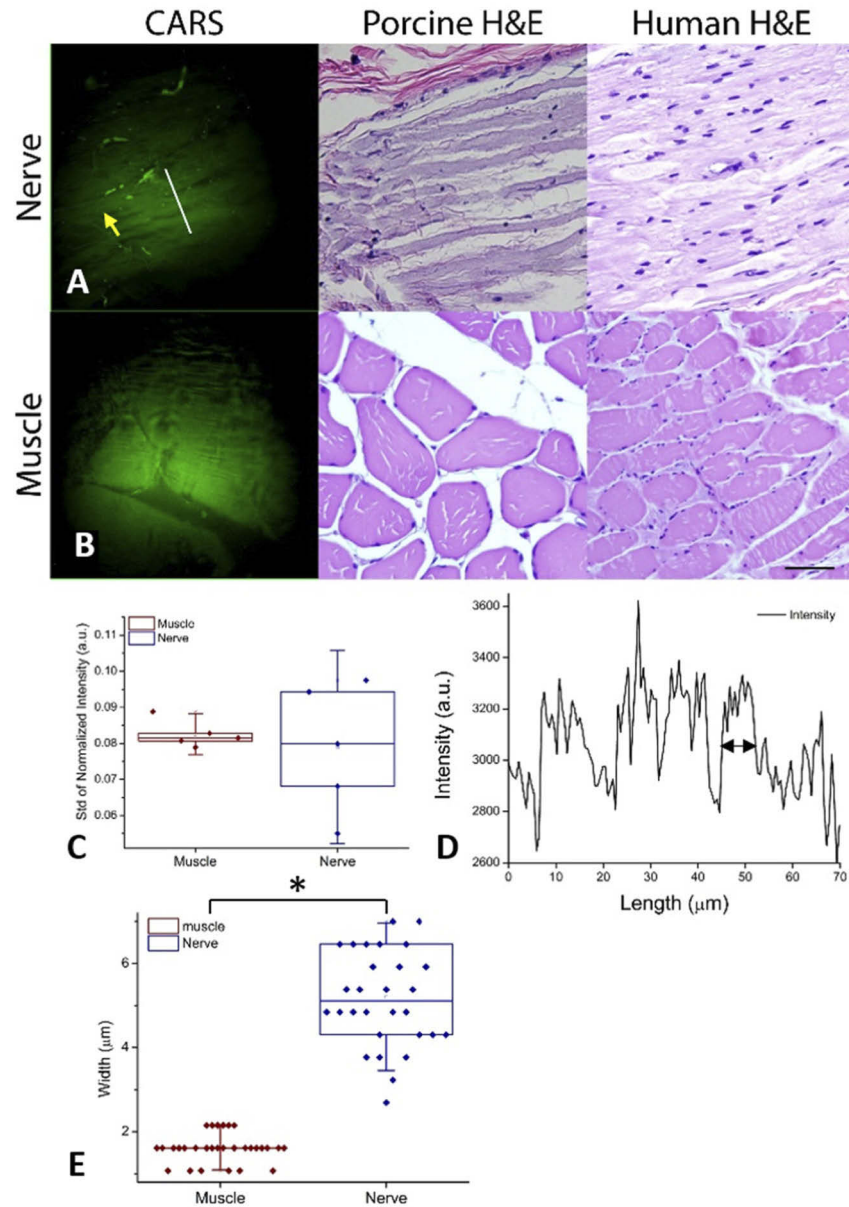
Fig. 3. A: CARS and H&E images of RLN. Straps with Schwann's nuclei (the darker regions depicted by the yellow arrow) can be seen in the CARS image similar to the H&E images. B: The muscle appears as longitudinal narrow and elongated fiber-like bands. Scale bar: 50 μm. C: Box plot depicting the mean (horizontal line in the box) and standard deviation (whiskers) of the values (standard deviation of normalized intensity pixels) representing each CARS image of RLN and muscle (P>0.05, using a paired student's t-test). D: Signal intensity of the white line marked in Fig. 2(A). An example of peak to peak measurement (width) is shown using the black double-sided arrow. E: Box plot representing the mean (horizontal line in the box) and standard deviation (whiskers) of the quantified widths of the bands/ straps. The dots represent the data distribution. Asterisk indicates a statistical significance of P<<0.01 using a paired student's t-test.

deviation of normalized intensity pixels) representing each of the CARS images of RLN and muscle, and the dots show the data distribution. The data from the muscle and RLN show an overlap, and the p-value from a paired student's t-test was greater than 0.05. This indicated that the CARS images of the RLN and muscle could not be differentiated using this method. However, since the width of the straps were different for the RLN and muscle, the width was used as the parameter for differentiation between the nerve and the muscle. Thirty regions of interest (an example is depicted by the white line in Fig. 3(A)) were randomly chosen for each tissue type and each of these selected regions had multiple straps/bands. The signal intensity was plotted for each of these regions as shown in Fig. 3(D). Since the bands were the brighter regions in both tissues, the peak to peak distances from the signal intensity plots were calculated and averaged to be the width of each band. This was done for all 30 regions of interest selected. The box plot in Fig. 3(E) shows the mean and standard deviation of the widths of these randomly selected straps/bands of RLN and muscle, and the dots show the actual data from each sample (width of strap/band). A paired student's t-test was performed, showing a statistically significant difference ($P \ll 0.01$) between the tissue types, indicating the ability to successfully differentiate between the RLN and muscle.

3.2. iCARS imaging: automatic tissue identification and differentiation

3.2.1. Regular retrained inception-v3 neural networks

We selected Inception-v3 from many deep learning algorithms with pre-trained weights implemented in Keras based on the trade-off criteria between costs and performance. The size of input 3D CARS images of our models is set to (500, 500, 3). The to-train part is the average 2D pooling layer + flatten layer + 6-dimensional output 'Softmax' dense layer. Thus, the output can be considered as the probability for each class while the sum of those 6 classes' probabilities reasonably equals to 1 and the loss function is set as the categorical cross-entropy.

3.2.2. Data preprocessing

For precision, the training data and testing data need to be randomly selected from the very beginning at the tissue level. Furthermore, the number of images in each class during training must be equal. Therefore, after rotating (x4), taking the inverse (x2) and cropping (x4), the number of images from each class was increased (nerve: from 40 to 1,280; fat: from 64 to 2,048; muscle: from 87 to 2,784; lymph node: from 46 to 1,472; thyroid: from 429 to 13,728; parathyroid: from 484 to 15,488). Then, 1,280 images were randomly chosen from each class for the data balance. Our training dataset contains 5,568 images (928 images for each class) and testing dataset contains 2,112 images (352 images for each class). Table 2 shows the procedure of increasing the number of datasets.

Table 2. The original data is the raw data scanned from machine. After rotation, inverse and crop data operations, the total dataset was increased to 36,800 images of size of 500 × 500 × 3.

	Nerve	Fat	Muscle	Lymph node	Thyroid	Parathyroid	Total
Original	40	64	87	46	429	484	1150
Rotate 90°(x4)	160	256	348	184	1,716	1936	4600
Inverse (x2)	640	1024	696	368	3,432	3,872	9200
Crop 500 × 500 (x4)	1,280	2,048	2,784	1,472	13,728	15,488	36,800

3.2.3. Training models

All training programs were run on a standard p2.xlarge instance of Amazon web services. The models were compiled with Adam optimizer (learning rate = $1e^{-4}$) and categorical cross-entropy loss function. The batch size of each training step was 16 and the epoch number was 10. At the end of each training step, 100 testing samples randomly chosen from the testing dataset were used as a step-by-step validation to guarantee that models converge in reality.

3.2.4. Model performance

The regular retrained Inception-v3 model achieved an accuracy of only 80.21% on the testing dataset (depicted in Fig. 4).
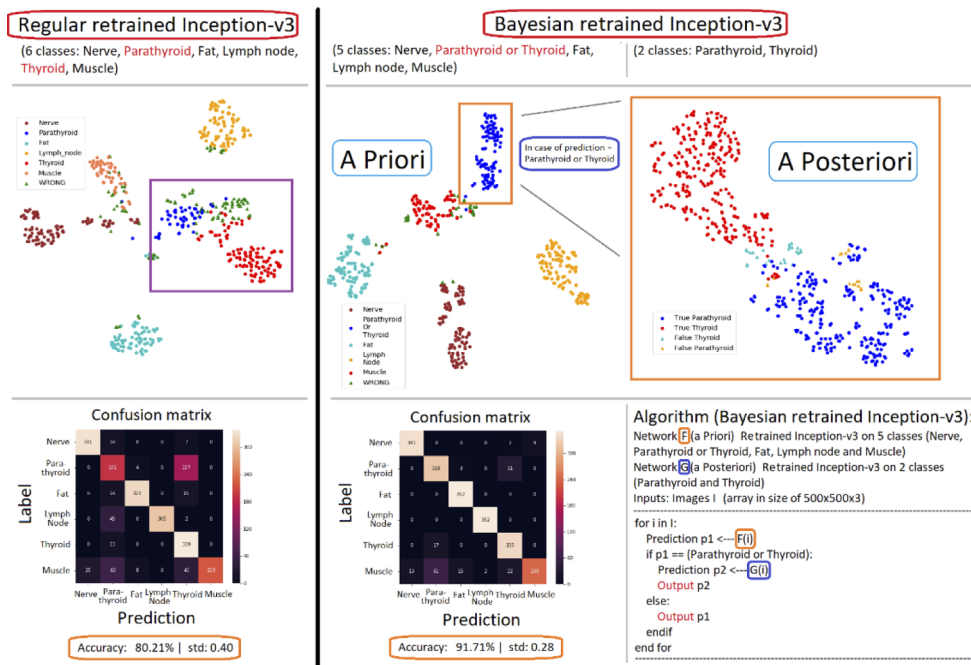
Fig. 4. Performance comparisons between regular retrained Inception-v3 and Bayesian retrained Inception-v3 models. The accuracy of the network has been increased from 80.21% to 91.71% with the use of the Bayesian retrained Inception-v3 model.

This could be due to the presence of different types of tissues at the same time in one image, sometimes even overlapped. In such situations, the model could predict the wrong tissue type, such as in Fig. 5(A). Meanwhile, there are also issues with tissue structure itself that we need to address.

In Fig. 5(B), for example, parathyroid and thyroid tissues are similar in structure. Thus, the regular retrained Inception-v3 neural networks can frequently confuse thyroid with parathyroid glands. While using t-SNE 2D-visualization, there was a high similarity between parathyroid and thyroid. The confusion matrix in Fig. 4 also shows how the model confuses these two tissues. The regular retrained Inception-v3 model predicts the parathyroid tissues correctly half the time and incorrectly the other half. Mathematically, the high similarity between parathyroid and thyroid tissues could also disturb the predictions of other image types such as lymph nodes.
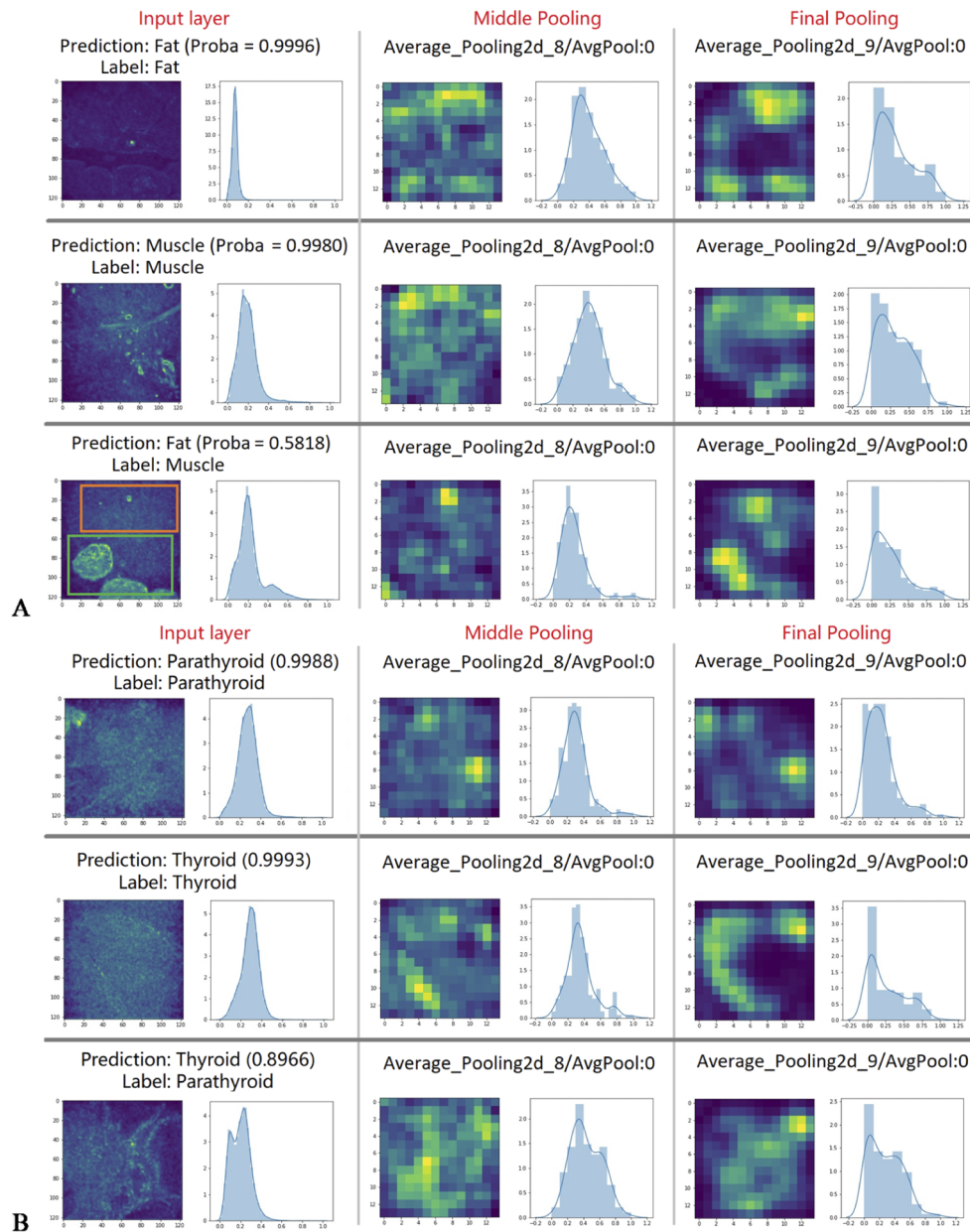
Fig. 5. Results from the regular retrained Inception-v3 model. A: Tissue types have been identified accurately in the first and second row. In some cases, such as in the 3rd row, the image contains half adipose (green rectangle) and half muscle (orange rectangle). The model has predicted adipose. However, the probability is very low (0.5818). B: Tissue types have been identified accurately in the first and second row. However, in the 3rd row, the image was misclassified as thyroid with a high probability (0.8966) while the label is parathyroid

3.2.5. Algorithm improvement

To resolve these problems, Bayesian retrained Inception-v3 neural networks were invented. First, we combined parathyroid and thyroid into one class, which led to a total of 5 classes: nerve, parathyroid/thyroid, adipose, lymph node, and muscle. This was called an *a priori* dataset. The model trained on this dataset provides prior probabilities of the tissue. Next, we selected parathyroid and thyroid as 2 separate classes (*a posteriori* dataset). To better compare the performance of the new model with the regular retrained model, the training dataset for the new model was picked from the regular retrained Inception-v3 model's training dataset. The same was done in the case of the testing dataset. Finally, we trained two regular retrained Inception-v3 models with the 5 classes and 2 classes, named as *a priori* retrained Inception-v3 neural networks and *a posteriori* retrained Inception-v3 neural networks, respectively. Once the two networks were trained, the algorithm shown in Fig. 3 was applied on the testing dataset.

As shown in Fig. 4, the combination of the two networks reduced the misclassifications (from 80.21% to 91.71%), particularly those between parathyroid and thyroid glands.

3.2.6. CARS imaging of thyroid cancer

Human thyroid cancer tissues were imaged using both CARS and histology and the results were compared. Figures 6 (A) and (B) depict the CARS and corresponding H&E images of normal human thyroid tissue. Figures 7 (A) and (B) depict the CARS and corresponding H&E images of cancerous human thyroid tissue. A high similarity was observed between the CARS images and the corresponding histological images. These results are in accordance with histological images found in literature of human thyroid cancer [48].
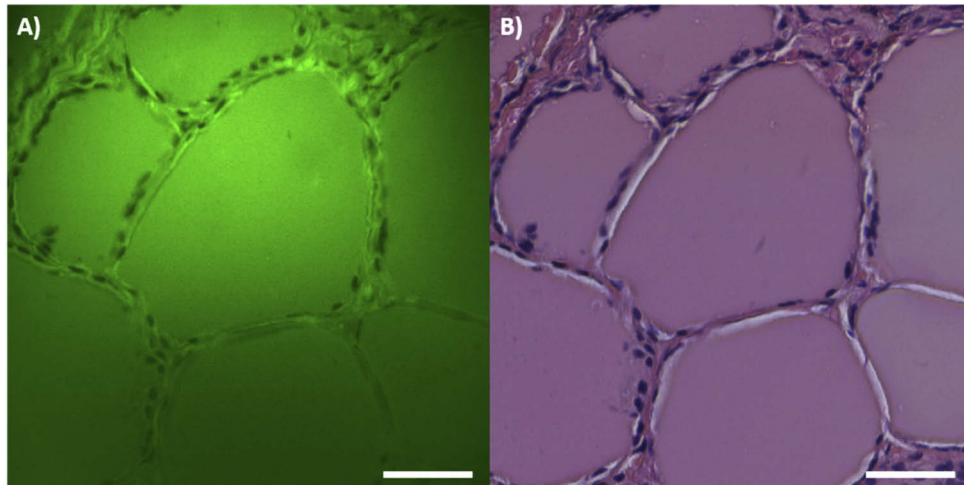

Fig. 6. A: CARS image of normal human thyroid tissue. B: Corresponding H&E image of normal human thyroid tissue. Scale bars: 50 μm

Figures 8 (A) and (B) depict the CARS and corresponding histological images of an area consisting of both normal and cancerous human thyroid tissue. A distinct boundary between the normal and cancerous tissue can be seen in both the images. This shows the potential of CARS imaging to not only differentiate between normal and cancerous tissue, but also detect clear margins.

The deep learning algorithms developed were used to differentiate between the normal and cancerous thyroid tissue, and to detect the cancer margins. This procedure is described in the next section.
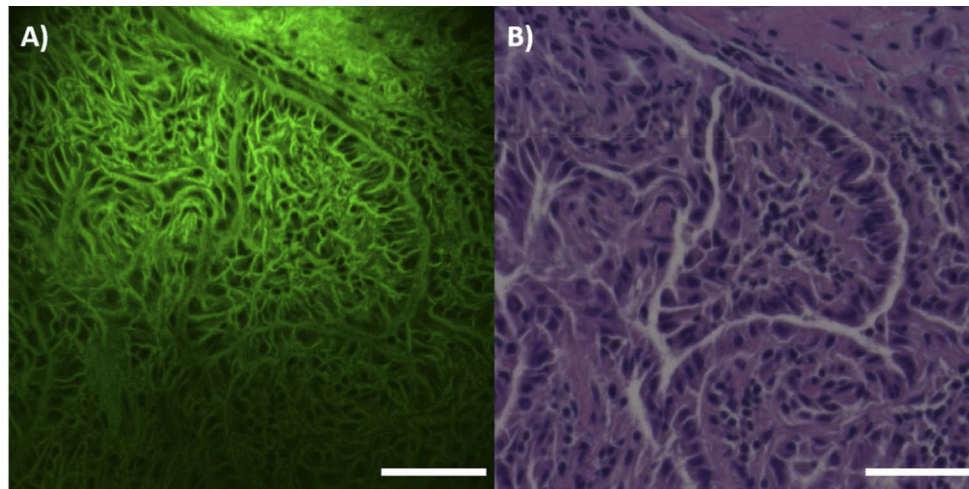
Fig. 7. A: CARS image of cancerous human thyroid tissue. B: Corresponding H&E image of cancerous human thyroid tissue. Scale bars: 50 μm
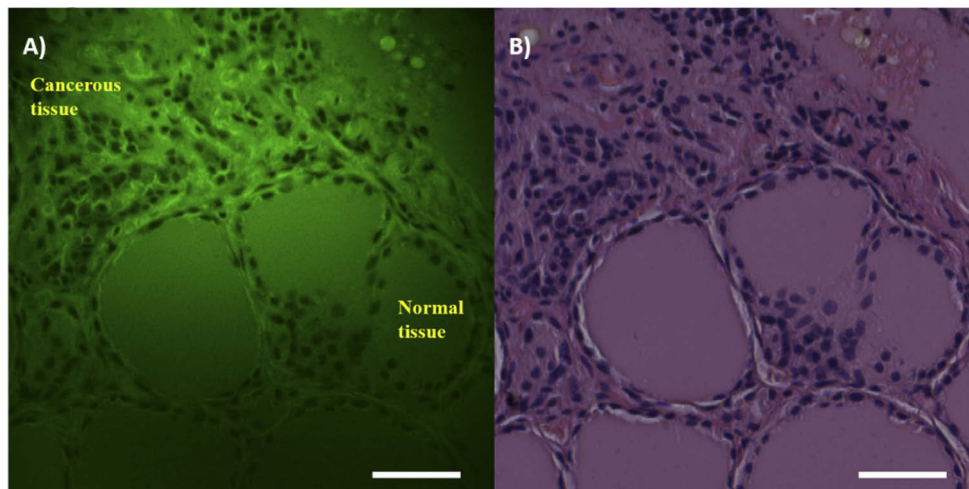

Fig. 8. A: CARS image of a region consisting of both normal and cancerous human thyroid tissue. B: Corresponding H&E image. Scale bars: 50 μm

3.2.7. Automatic thyroid cancer margin detection

To build the detection model, we needed the correct classifications of cancerous and normal tissues. For this purpose, we first manually segment the region of thyroid cancer on the eight hundred ninety-five images (of size of 512 pixels by 512 pixels). The images with the manually segmented area were then used as our original dataset. Before augmentation by rotation and inversion (x8), 713 images were randomly chosen as the training dataset, and the remaining 182 were the testing dataset. After data augmentation, there were a total of 5,704 images for training and 1,456 images for testing. The training and testing parts did not overlap with each other.

The U-Net structure was selected as the mainframe for our model based on its excellent accuracy and high efficiency (Fig. 9). The U-Net consists of two parts: the downsampling layer that extracts features from the input image, and the upsampling structure that reconstructs the segmentation image step by step. The U-Net is a symmetric structure and has the same number of layer blocks for downsampling and upsampling. The convolutional layers' kernel size is identical, 3×3. The padding within a block is 1, and the channel is fixed. The padding between blocks is set as two, and the channel doubles to ensure that downsampling blocks extract features step-by-step. To avoid over-fitting, the dropout layer with a 50% ratio was added to the last layer of the last downsampling block. The upsampling part starts from the last layer of the deepest blocks (dropout layer).
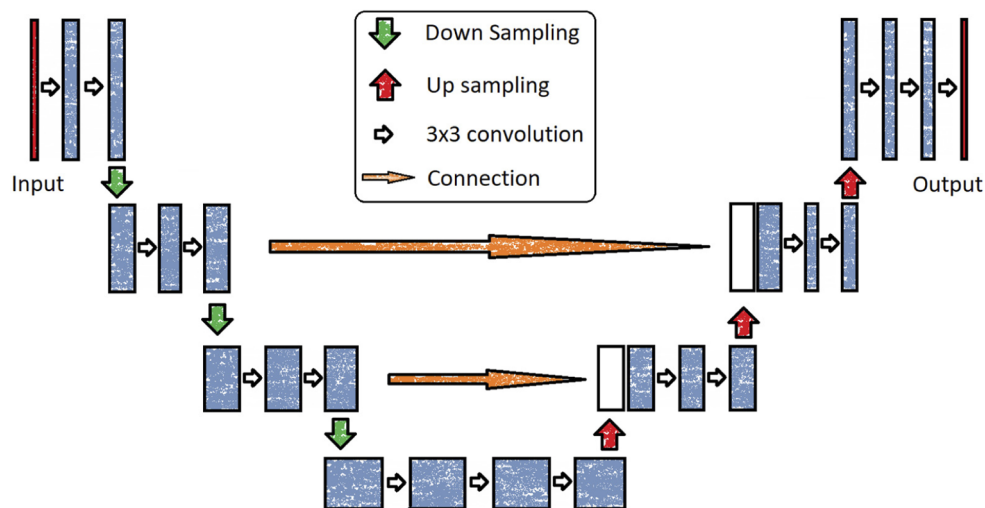

Fig. 9. The structure of U-Net. The skip connection between the first block and the last block was removed to decrease the input image restrictions. The Batch-normalization layers are not shown here. The input and output images size are 512×512. The activations were set as 'Relu' while the last layer was using 'sigmoid.'

Similar to the downsampling part, the padding within a block was set as 1, and the channel of layers stays the same while the padding between blocks is two, and the channel is divided into 2. The skip connections are set between corresponding downsampling and upsampling blocks to help the reconstruction. However, the skip connection between the input block and the output block is canceled because of the robustness. With this structure, the output image size is identical to the input image; therefore, each pixel on the image was classified as either cancerous or normal tissue after the segmentation procedure. The batch normalization layers were set at the start of all the blocks (except the input block and the output block) to help the model converge, not shown in Fig. 9 below. The last layer predicts the probability of being cancer or not for each pixel in the input image. Therefore, the binary cross-entropy is the best option
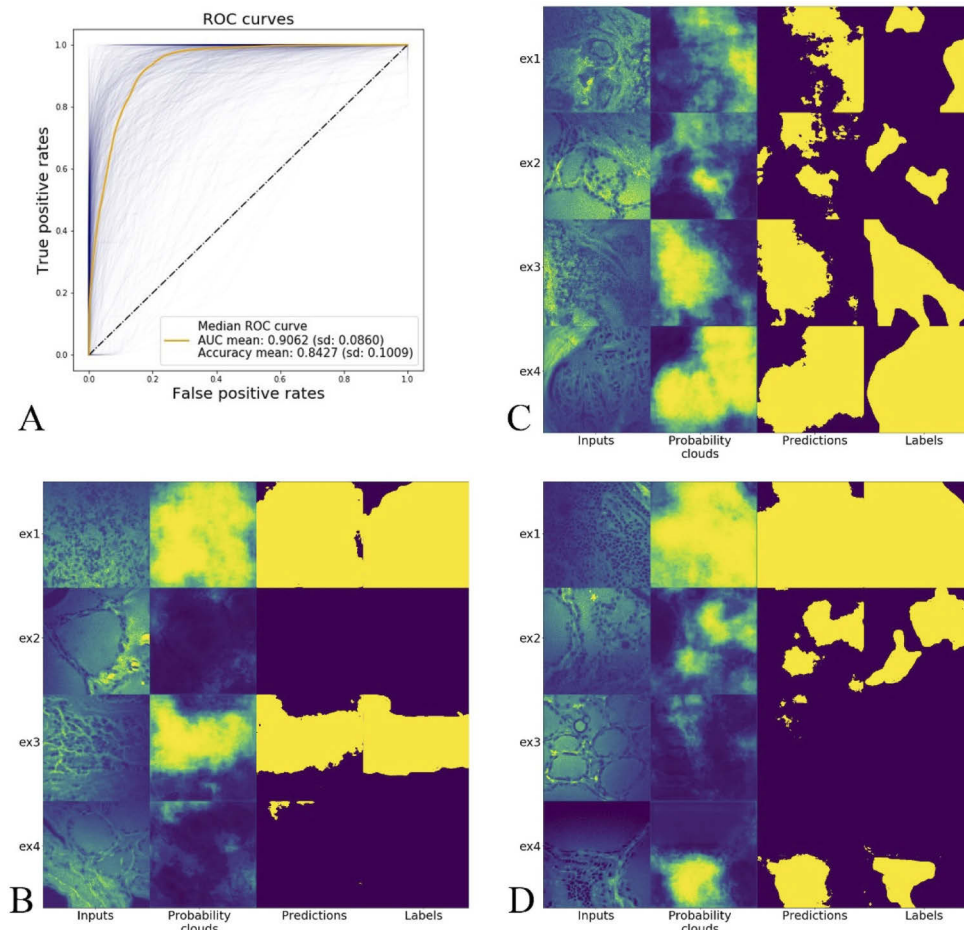
Fig. 10. Results of the automated cancer margin detection. A: The performance of the U-Net segmentation neural networks. The blue curves are ROC curves of every validation or test image. The orange curve is the median ROC curves among those curves. B, C, and D: Each horizontal line depicts four samples from the test dataset. The 'probability clouds' are the probabilities predicted by U-net for every image pixel. The yellow part of the column 'Predictions' depict the probability of cancer is higher than 50%. The last column is area of cancerous tissue segmented using our manual method for comparison.

for loss function because the segmentation becomes a two-class classification for every pixel of the input image. The model was tested every ten training batch rounds on testing data set to supervise the convergence during the training procedure by loss values. The test during training procedure was applied with ten randomly picked-up images from the test dataset.

We use Adam as the optimizer and $1e^{-4}$ as the learning rate. Batch size is set as 32 images and epoch as 2,000. All training and testing tasks were performed on the server of p2-xlarge with GPUs of Amazon computing cloud. After accomplishing the training procedure (the binary cross-entropy variation is less than $1e^{-4}$), the U-Net model was tested on the whole test dataset. The accuracy was calculated, and receiver operating characteristic (ROC) curve was plotted for every test image. In Fig. 10 (A), every blue line represents a ROC curve of a testing image. The mean area under the ROC Curve (AUC) (0.9062 ± 0.0860) and the mean accuracy (0.8427 ± 0.1009) values were calculated. The orange line in the image is the ROC curve with a

median AUC value. Figures 10 (B), (C), and (D) describe the process of tissue identification and labeling. Each of these figures contains four samples from the test dataset labeled ex1, ex2, ex3, and ex4 (rows). The first column of each figure depicts the input image. The second column shows the probability cloud, which is the probability predicted by the U-Net for every pixel in the input image. The lighter the pixel, the higher the probability of cancer for the pixel. The yellow part of the third column titled "Predictions" are the pixels where the probability of cancer was higher than 50%. The last column depicts the areas (of cancerous tissue) segmented using our manual method for comparison.

3.3.   CARS imaging of mouse thyroid in situ

The mouse thyroid and surrounding tissues were imaged *in situ* to test the potential of this technique as an intraoperative guiding tool. Figure 11 (A) shows the region imaged by CARS and CARS images of the mouse thyroid and fat. The image at the center of Fig. 11 (B) shows the 3D image of murine thyroid. Figures 11 B (a) to (i) depict transverse cross-sectional images at different depths. This shows that CARS imaging can be used to obtain images along the depth of the sample, without losing resolution. Although the thyroid tissue was not visible at the surface (Fig. 11 B (a)), it could be clearly identified from the images obtained deeper in the tissue (Fig. 11 B (c) onwards) and the reconstructed 3D image. This showed that the tissues could be identified by CARS imaging without any resection, which can significantly reduce the time of the identification procedure. This also showed that the iCARS imaging system implemented in this study can image more than 200 µm below the surface of tissues.
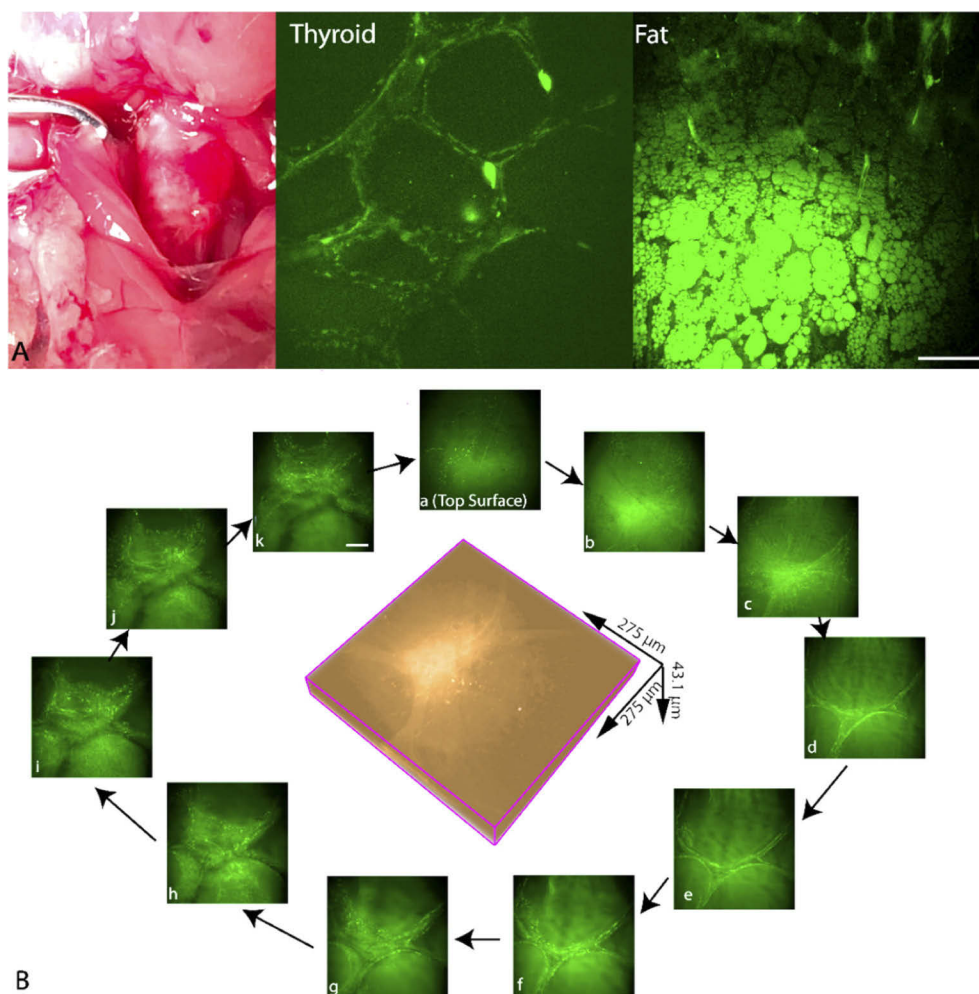
Fig. 11. A: CARS images of mouse thyroid and fat in situ. B: 3D image of murine porcine thyroid. (a) to (i) are the transverse slices at different depths of 0 (top surface), 2.2 μm, 5.7 μm, 10.7 μm, 15.8 μm, 19 μm, 24.6 μm, 29.8 μm, 34.5 μm, 38.5 μm, and 43.1 μm, respectively. Scale bar: 50 μm.

4. Discussion

Identification of normal parathyroid glands and RLN during thyroid surgery is critical to prevent unwanted dissection of these two vital tissues. Dissection of these tissues could cause permanent postoperative hypoparathyroidism and vocal cord paralysis (7, 8). Imaging methods including the use of labels or exogenous agents pose significant limitations and hurdles for clinical use. Thus, several studies have focused on using label-free imaging to identify the parathyroid glands (12, 22). Conti de Freitas et al. successfully identified parathyroid glands using Optical Coherence Tomography (OCT). However, due to the low spatial resolution and low contrast of OCT imaging, only coarse grained structural images were analyzed, losing all cellular level information. The identification of RLN was also not discussed (12). Thus, OCT systems available in the market do not address the issues of unwanted dissection of normal parathyroid glands and RLN. Hence there is a need for a more sensitive, high resolution, label-free imaging method to identify the parathyroid and RLN.

We first analyzed porcine tissues in this study since they are similar to humans. We demonstrated that the iCARS imaging method is capable of providing useful quantitative microscopic information of tissues that is critical during thyroid surgery, especially for thyroid carcinoma. Normal parathyroid glands appeared distinctive in CARS images, with the characteristics of gathered round shape, and small and similar sized nuclei evenly distributed inside pool-like structures. The other tissues appeared entirely different in CARS images. For example, thyroid had a spherical follicle structure with homogeneous internal colloid, and dark nuclei distributed along the boundary of the spherical structure. The lymph node had size and shape varied nuclei gathered and embedded in net-like fibrous connective tissues, while the adipose showed up as a bright area with a smooth edge. Due to these distinct characteristics of all the tissues surrounding the parathyroid, CARS imaging helped differentiate the parathyroid gland from the surrounding tissues.

However, certain similarities between these tissues may reduce the accuracy in differentiating the parathyroid from thyroid and lymph nodes. For example, both the parathyroid and lymph nodes have gathered nuclei distribution while the parathyroid and thyroid images include follicle structures. Therefore, additional analysis was performed to achieve the optimal accuracy for identification of parathyroid. By using the intensity and width analysis, all tissues could be differentiated appropriately.

From classification to segmentation, this model can be precise with sufficient well-labeled data. For classification, we replaced the regular retrained Inception-v3 model with the Bayesian retrained Inception-v3 due to high similarity between parathyroid and thyroid images, which caused misidentifications. Another reason for this change was that the volume of parathyroid and thyroid datasets was much larger than other types of tissues. Therefore, the Bayesian retrained Inception v3 was more apt. The *a posteriori* Inception-v3 sub-model of Bayesian retrained Inception-v3 was more accurate with the parathyroid-thyroid classification than the regular retrained Inception-v3, which helped the Bayesian retrained Inception-v3 to decrease misclassifications. For segmentation, due to the similarities between the parathyroid and thyroid, which are located in close proximity to each other, an auto-segmentation tool can help the surgeons better. The U-Net structure was trained using the manually labeled parathyroid and thyroid images and thus achieved high accuracy in segmenting the tissues. As the types of tissues to classify and segment increase, a systematic segmentation model which allows quick identification of particular tissue units and its location would be more convenient to surgeons. This would be a part of our future work.

To the best of our knowledge, this is the first reported study of integrating label-free CARS imaging and deep learning to identify complex, heterogeneous tissues such as parathyroid adenoma, hyperplasia, and surrounding neck tissues and nerves in real time. Using iCARS imaging as an intraoperative tool that performs optical biopsies to distinguish parathyroid adenoma

from hyperplasia will significantly reduce the operation time. This system can identify the hyperplasia tissue instantly during parathyroid hyperplasia surgeries and identify the parathyroid gland and RLN during thyroid surgeries, which in turn will drastically lower the rate of damage to these two critical tissues. Total thyroidectomy followed by central neck dissection is a typical treatment for papillary thyroid carcinoma with suspicion for central lymph node metastasis. This iCARS system can provide guidance by differentiating neck tissues in real-time when the tissues cannot be clearly identified by gross visualization. This could help with only necessary tissues being dissected, and the others preserved. If the surgeon suspects that the parathyroid gland was accidentally resected during surgery, an immediate decision can be made aided by iCARS imaging, instead of waiting for more than thirty minutes to confirm using frozen section method. Once the delay is minimized, the function of the resected parathyroid gland can be maximally preserved after auto-transplantation. Applying iCARS imaging on remaining parathyroid glands, we can separate parathyroid adenoma from hyperplasia. iCARS helps with better decision-making throughout the operation.

As a part of our future work, we are also developing a handheld microscope probe to replace the fixed microscopy stand for intraoperative evaluation. Previously, we have designed a small microendoscope probe for CARS imaging for robotic-assisted radical prostatectomy by inserting it with a robotic arm [49]. However, thyroidectomies are mostly open surgeries and less dimensionally challenged by the inner dimension of a surgical robot arm. Hence, the size of the iCARS microendoscope will be an order larger and more sensitive than a robotic-oriented CARS microendoscope. We are also developing an all-fiber based high-speed CARS system that is sensitive to C-H bonds. The price of a fixed-wavelength high-speed fiber laser is now affordable such that low-cost iCARS imaging systems can improve the mobility and stability significantly without sacrificing the performance.

5. Conclusions

This study demonstrates the successful design and application of high resolution, label-free iCARS imaging to identify subtle yet complex patterns of various neck tissues encountered during sensitive thyroid surgery. This could potentially diminish the two most common risks of thyroid surgery, namely, damage to nerves controlling the vocal cords and damage to parathyroid glands that control calcium homeostasis. The iCARS imaging method can be used to identify parathyroid adenoma and hyperplasia, and differentiate them from the thyroid glands, lymph nodes, and adipose tissues morphologically without staining. Its advanced deep learning model improves the accuracy and realizes full automation of tissue identification on the spot for clinical use. In this study, pig tissues were harvested in fresh, imaged, and analyzed by the iCARS imaging method immediately to show the performance of imaging and differentiation while the *in situ* imaging of mouse thyroid and the porcine thyroid tissues demonstrated the potential use of the iCARS imaging method to identify parathyroid glands, diagnose thyroid cancer, and detect cancer margins. The results indicate that iCARS imaging can be used as a new modality for on-the-spot decision-making during parathyroid and thyroid surgery, filling the current need of high resolution molecular and cellular imaging strategies during surgical procedures without using exogenous agents. Studies are underway to develop a handheld microscope probe to replace the current fixed microscopy system for intraoperative procedures. We are also building a complete fiber-based system to improve the mobility and stability of the system. Since cancer cells and nerves are similar in different organs, high precision iCARS imaging, potentially, can be used to detect cancer margins or prevent nerve damage during surgeries, such as in prostate cancer surgery, where it is critical to prevent nerve impairment.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. G. H. Sun, S. DeMonner, and M. M. Davis, "Epidemiological and economic trends in inpatient and outpatient thyroidectomy in the United States, 1996-2006," Thyroid **23**(6), 727–733 (2013).
2. H. S. Kazaure and J. A. Sosa, "Surgical hypoparathyroidism," Endocrinology and Metabolism Clinics of North America **47**(4), 783–796 (2018).
3. S. M. Wilhelm, T. S. Wang, D. T. Ruan, J. A. Lee, S. L. Asa, Q. Y. Duh, G. M. Doherty, M. F. Herrera, J. L. Pasieka, N. D. Perrier, S. J. Silverberg, C. C. Solorzano, C. Sturgeon, M. E. Tublin, R. Udelsman, and S. E. Carty, "The American Association of Endocrine Surgeons Guidelines for Definitive Management of Primary Hyperparathyroidism," JAMA Surg. **151**(10), 959–968 (2016).
4. A. El-Husseini, K. Wang, A. Edon, D. Saxon, F. Lima, D. Sloan, and B. P. Sawaya, "Value of intraoperative parathyroid hormone assay during parathyroidectomy in dialysis and renal transplant patients with secondary and tertiary hyperparathyroidism," Nephron **138**(2), 119–128 (2018).
5. G. Trinh, S. I. Noureldine, J. O. Russell, N. Agrawal, M. Lopez, J. D. Prescott, M. A. Zeiger, and R. P. Tufano, "Characterizing the operative findings and utility of intraoperative parathyroid hormone (IOPTH) monitoring in patients with normal baseline IOPTH and normohormonal primary hyperparathyroidism," Surgery **161**(1), 78–86 (2017).
6. O. Edafe, E. E. Collins, C. S. Ubhi, and S. P. Balasubramanian, "Current predictive models do not accurately differentiate between single and multi gland disease in primary hyperparathyroidism: a retrospective cohort study of two endocrine surgery units," annals **100**(2), 140–145 (2018).
7. H. Oda, A. Miyauchi, Y. Ito, K. Yoshioka, A. Nakayama, H. Sasai, H. Masuoka, T. Yabuta, M. Fukushima, T. Higashiyama, M. Kihara, K. Kobayashi, and A. Miya, "Incidences of unfavorable events in the management of low-risk papillary microcarcinoma of the thyroid by active surveillance versus immediate surgery," Thyroid **26**(1), 150–155 (2016).
8. W. C. Tseng, Y. C. Pei, A. M. Wong, H. Y. Li, and T. J. Fang, "Distinct disease and functional characteristics of thyroid surgery-related vocal fold palsy," Thyroid **26**(7), 943–950 (2016).
9. S. M. Sadowski, P. Soardo, I. Leuchter, J. H. Robert, and F. Triponez, "Systematic use of recurrent laryngeal nerve neuromonitoring changes the operative strategy in planned bilateral thyroidectomy," Thyroid **23**(3), 329–333 (2013).
10. M. K. Applewhite, M. G. White, M. Xiong, J. D. Pasternak, L. Abdulrasool, L. Ogawa, I. Suh, J. E. Gosnell, E. L. Kaplan, Q. Y. Duh, P. Angelos, W. T. Shen, and R. H. Grogan, "Incidence, risk factors, and clinical outcomes of incidental parathyroidectomy during thyroid surgery," Ann Surg Oncol **23**(13), 4310–4315 (2016).
11. K. Huang, D. Luo, M. Huang, M. Long, X. Peng, and H. Li, "Protection of parathyroid function using carbon nanoparticles during thyroid surgery," Otolaryngol Head Neck Surg **149**(6), 845–850 (2013).
12. A. Watanabe, M. Taniguchi, Y. Kimura, S. Ito, M. Hosokawa, and S. Sasaki, "Efficient, effective, safe procedure to identify nonrecurrent inferior laryngeal nerve during thyroid surgery," Head Neck **38**(4), 573–577 (2016).
13. Z. Al-Qurayshi, G. W. Randolph, M. Alshehri, and E. Kandil, "Analysis of variations in the use of intraoperative nerve monitoring in thyroid surgery," JAMA Otolaryngol Head Neck Surg **142**(6), 584–589 (2016).
14. F. Y. Chiang, I. C. Lu, H. C. Chen, H. Y. Chen, C. J. Tsai, P. J. Hsiao, K. W. Lee, and C. W. Wu, "Anatomical variations of recurrent laryngeal nerve during thyroid surgery: how to identify and handle the variations with intraoperative neuromonitoring," The Kaohsiung Journal of Medical Sciences **26**(11), 575–583 (2010).
15. M. Long, D. Luo, F. Diao, M. Huang, K. Huang, X. Peng, S. Lin, and H. Li, "A carbon nanoparticle lymphatic tracer protected parathyroid glands during radical thyroidectomy for papillary thyroid non-microcarcinoma," Surg Innov **24**(1), 29–34 (2017).
16. H. Kai, L. Xixia, L. Miaoyun, C. Qinchang, P. Xinzhi, L. Dingyuan, and L. Honghao, "Intraoperative nerve monitoring reduces recurrent laryngeal nerve injury in geriatric patients undergoing thyroid surgery," Acta Oto-Laryngologica **137**(12), 1275–1280 (2017).
17. J. A. Sosa, J. W. Hanna, K. A. Robinson, and R. B. Lanman, "Increases in thyroid nodule fine-needle aspirations, operations, and diagnoses of thyroid cancer in the United States," Surgery **154**(6), 1420–1427 (2013).
18. L. C. Conti de FreitasC, E. Phelan, L. Liu, J. Gardecki, E. Namati, W. C. Warger, G. J. Tearney, and G. W. Randolph, "Optical coherence tomography imaging during thyroid and parathyroid surgery: a novel system of tissue identification and differentiation to obviate tissue resection and frozen section," Head Neck **36**, 1329–1334 (2014).
19. X. H. Bian, S. J. Li, L. Zhou, C. H. Zhang, G. Zhang, Y. T. Fu, and H. Sun, "Applicability of rapid intraoperative parathyroid hormone assay through fine needle aspiration to identify parathyroid tissue in thyroid surgery," Exp. Ther. Med. **12**(6), 4072–4076 (2016).
20. H. Hyun, M. H. Park, E. A. Owens, H. Wada, M. Henary, H. J. Handgraaf, A. L. Vahrmeijer, J. V. Frangioni, and H. S. Choi, "Structure-inherent targeting of near-infrared fluorophores for parathyroid and thyroid gland imaging," Nat Med **21**(2), 192–197 (2015).
21. Y. J. Suh, J. Y. Choi, Y. J. Chai, H. Kwon, J. W. Woo, S. J. Kim, K. H. Kim, K. E. Lee, Y. T. Lim, and Y. K. Youn, "Indocyanine green as a near-infrared fluorescent agent for identifying parathyroid glands during thyroid surgery in dogs," Surg Endosc **29**(9), 2811–2817 (2015).

22. Y. Shinden, A. Nakajo, H. Arima, K. Tanoue, M. Hirata, Y. Kijima, K. Maemura, and S. Natsugoe, "Intraoperative identification of the parathyroid gland with a fluorescence detection system," World J. Surg. **41**(6), 1506–1512 (2017).

23. M. A. McWade, M. E. Sanders, J. T. Broome, C. C. Solorzano, and A. Mahadevan-Jansen, "Establishing the clinical utility of autofluorescence spectroscopy for parathyroid detection," Surgery **159**(1), 193–203 (2016).

24. Q. R. Tummers, A. Schepers, J. F. Hamming, J. Kievit, J. V. Frangioni, C. J. van de Velde, and A. L. Vahrmeijer, "Intraoperative guidance in parathyroid surgery using near-infrared fluorescence imaging and low-dose Methylene Blue," Surgery **158**(5), 1323–1330 (2015).

25. M. A. McWade, C. Paras, L. M. White, J. E. Phay, A. Mahadevan-Jansen, and J. T. Broome, "A novel optical approach to intraoperative detection of parathyroid glands," Surgery **154**(6), 1371–1377 (2013).

26. M. A. McWade, C. Paras, L. M. White, J. E. Phay, C. C. Solorzano, J. T. Broome, and A. Mahadevan-Jansen, "Label-free intraoperative parathyroid localization with near-infrared autofluorescence imaging," The Journal of Clinical Endocrinology & Metabolism **99**(12), 4574–4580 (2014).

27. J. Falco, F. Dip, P. Quadri, M. de la Fuente, M. Prunello, and R. J. Rosenthal, "Increased identification of parathyroid glands using near infrared light during thyroid and parathyroid surgery," Surgical endoscopy **31**, 3737 (2017).

28. F. Hou, Y. Yu, and Y. Liang, "Automatic identification of parathyroid in optical coherence tomography images," Lasers Surg. Med. **49**(3), 305–311 (2017).

29. L. Gao, H. Zhou, M. J. Thrall, F. Li, Y. Yang, Z. Wang, P. Luo, K. K. Wong, G. S. Palapattu, and S. T. Wong, "Label-free high-resolution imaging of prostate glands and cavernous nerves using coherent anti-Stokes Raman scattering microscopy," Biomed. Opt. Express **2**(4), 915–926 (2011).

30. L. Gao, Z. Wang, F. Li, A. A. Hammoudi, M. J. Thrall, P. T. Cagle, and S. T. Wong, "Differential diagnosis of lung carcinoma with coherent anti-Stokes Raman scattering imaging," Arch. Pathol. Lab. Med. **136**(12), 1502–1510 (2012).

31. O. Uckermann, R. Galli, S. Tamosaityte, E. Leipnitz, K. D. Geiger, G. Schackert, E. Koch, G. Steiner, and M. Kirsch, "Label-free delineation of brain tumors by coherent anti-Stokes Raman scattering microscopy in an orthotopic mouse model and human glioblastoma," PloS one 9(9), e107115 (2014).

32. Y. Yang, F. Li, L. Gao, Z. Wang, M. J. Thrall, S. S. Shen, K. K. Wong, and S. T. Wong, "Differential diagnosis of breast cancer using quantitative, label-free and molecular vibrational imaging," Biomed. Opt. Express **2**(8), 2160–2174 (2011).

33. R. Galli, V. Sablinskas, D. Dasevicius, A. Laurinavicius, F. Jankevicius, E. Koch, and G. Steiner, "Non-linear optical microscopy of kidney tumours," J. Biophotonics **7**(1-2), 23–27 (2014).

34. F. B. Legesse, A. Medyukhina, S. Heuke, and J. Popp, "Texture analysis and classification in coherent anti-Stokes Raman scattering (CARS) microscopy images for automated detection of skin cancer," Computerized Medical Imaging and Graphics **43**, 36–43 (2015).

35. H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," Future Computing and Informatics Journal **3**(1), 68–71 (2018).

36. J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical image classification using synergic deep learning," Med. Image Anal. **54**, 10–19 (2019).

37. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016), 2818–2826.

38. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 (2014).

39. L. D. Nguyen, D. Lin, Z. Lin, and J. Cao, "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, (IEEE, 2018), 1–5.

40. M. Nishio, O. Sugiyama, M. Yakami, S. Ueno, T. Kubo, T. Kuroda, and K. Togashi, "Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning," PLoS One **13**, e0200721 (2018).

41. C. Haarburger, P. Langenberg, D. Truhn, H. Schneider, J. Thüring, S. Schrading, C. K. Kuhl, and D. Merhof, "Transfer learning for breast cancer malignancy classification based on dynamic contrast-enhanced MR images," in Bildverarbeitung für die Medizin 2018 (Springer, 2018), pp. 216–221.

42. M. Talo, U. B. Baloglu, Ö. Yıldırım, and U. R. Acharya, "Application of deep transfer learning for automated brain abnormality classification using MR images," Cognitive Systems Research **54**, 176–188 (2019).

43. J. Y. Lee, S.-H. Kim, D. W. Moon, and E. S. Lee, "Three-color multiplex CARS for fast imaging and microspectroscopy in the entire CHn stretching vibrational region," Opt. Express **17**(25), 22281–22295 (2009).

44. T. He, Z. Xue, and S. T. Wong, "A three-dimensional medical image segmentation app using graphic theory," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics* (BHI, 2016), 268–271.

45. Z. Xiangjun, X. Zhong, G. Xin, Z. Yisheng, and S. T. C. Wong, "VOLES: Vascularity-Oriented LEvel Set algorithm for pulmonary vessel segmentation in image guided intervention therapy," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009), 1247–1250.

46. S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," J. Comput. Phys. **79**(1), 12–49 (1988).

47. R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," IEEE Trans. Pattern Anal. Machine Intell. **17**(2), 158–175 (1995).

48. S. G. Silverberg, R. V. Hutter, and F. W. Foote Jr., "Fatal carcinoma of the thyroid: histology, metastases, and causes of death," Cancer **25**(4), 792–802 (1970).

49. X. Chen, X. Xu, D. T. McCormick, K. Wong, and S. T. Wong, "Multimodal nonlinear endo-microscopy probe design for high resolution, label-free intraoperative imaging," Biomed. Opt. Express **6**(7), 2283–2293 (2015).